APPLOVIN

Text Mining Analysis of Wordscapes Reviews

Written by: Abdul Bishar

Date: March 5th, 2020

Introduction and context:

After my first conversation with Helen Wu Merlin (Sr Director of Partnerships), I was delighted to discover that we happened to both love playing Wordscapes. Wordscapes is a popular word puzzle game developed by Texas based game studio, PeopleFun and has been downloaded over 10 million times. The current ratings for the game are 4.9 and 4.7 across android play store and apple store respectively.

After my conversation with Helen, I was inspired to take on a text mining exercise whereby I would web scrape over reviews for the game to answer questions such as

- What are some of the most common words people use to describe the game?
- What are people's general sentiment toward the game?
- What other valuable insights can be found by analyzing reviews for the game?

Metadata:

The data was pulled from Google Play Store on 03/05/2020.

The reviews were taken from 25/12/2019 – 03/05/2020

Number of unique reviews in dataset = 2569 (Total reviews in app store is 759,000)

NB: All analysis and visualizations produced in R Studio

Analysis:

While the number of reviews analyzed in this exercise represent a very small sample of the total reviews on the app store (0.3%), I believe that there is still value to be found in 2569 reviews sampled. Moreover, these are the most current reviews over the last 2 months or so and therefore should hold more weight when analyzing people's opinions due to its recency

The first step in the analysis involved reading in the data and removing common stop words such as and,or,I,we, he, she etc. This left a total of 8518 words to analyze

Next, I took a look at the most frequent words used in the reviews and visualized it in the form of a bar graph as well as a word cloud.

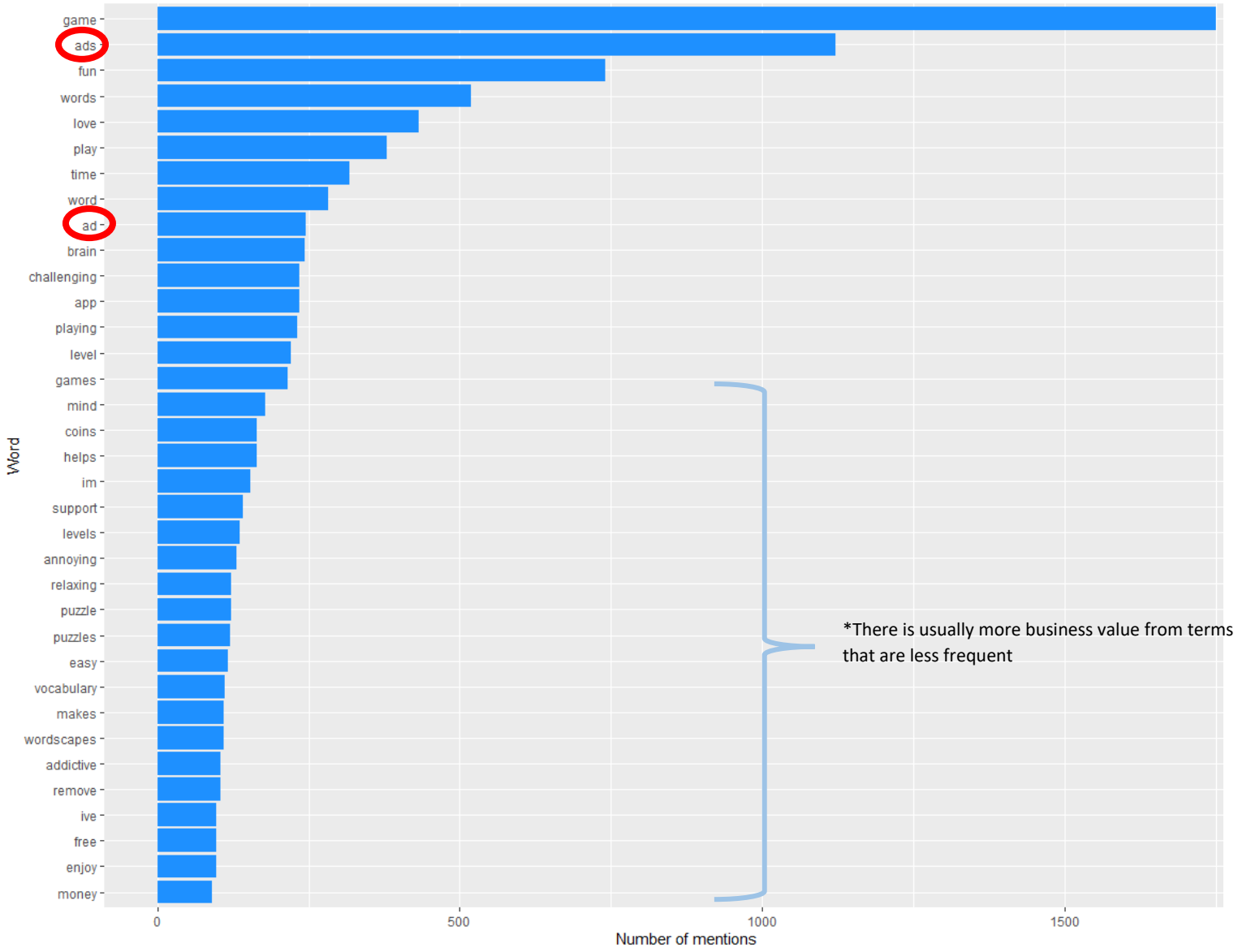

Figure 1: Frequency diagram

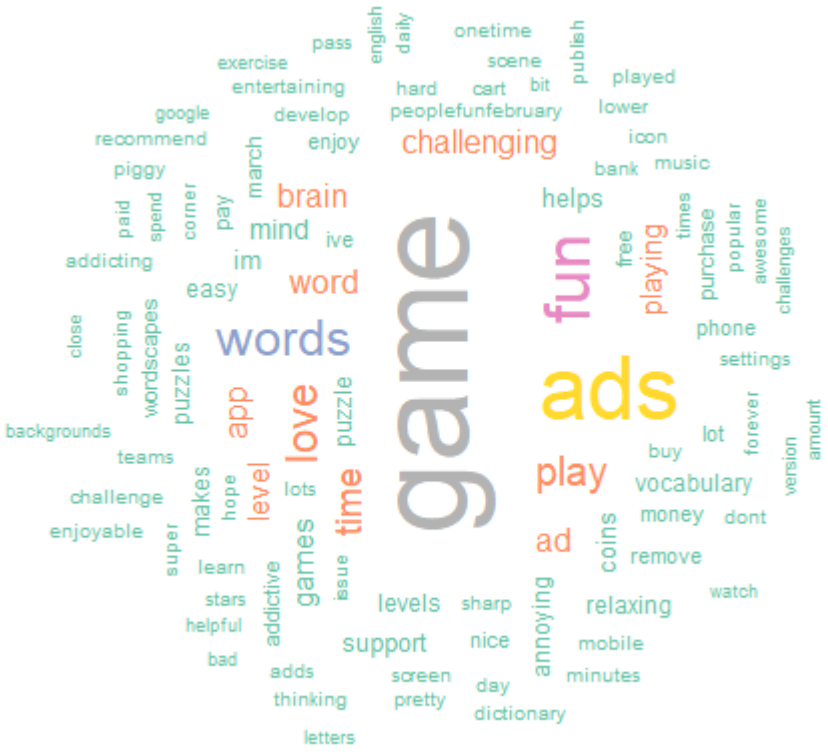

Figure 2:Frequency Word Cloud

One of the first things we notice is that the words “Ads” and “Ad” are mentioned quite frequently, 1121 and 246 times respectively. Clearly people are talking about the ads they are experiencing in the game and so I marked it for further investigation

Of course, there are other takeaways here. We can see words like “Mind”, “Brain”, “Challenge”,”Love” and “Addictive” which are positive words we would like to see popping up frequently. I did however also notice the word “Annoying” pop up frequently.

Besides taking a look at single word frequencies. We can also take a look at Bigrams frequencies i.e. the most common word pairs that appear in the reviews.

The most frequent Bigrams are shown in the table below

	word1	word2	n
1	love	game	156
2	fun	game	140
3	remove	ads	91
4	hope	helps	61
5	ads	support	60
6	develop	publish	60
7	mobile	games	60
8	popular	mobile	60
9	publish	wordscapes	60
10	support	teams	60
11	teams	develop	60
12	ads	forever	59
13	cart	icon	59
14	corner	game	59
15	game	scene	59
16	icon	lower	59
17	onetime	purchase	59
18	shopping	cart	59
19	wordscapes	hope	59
20	word	game	54
21	pass	time	49
22	piggy	bank	48
23	game	fun	45
24	playing	game	44
25	fun	challenging	42
26	fun	play	36

Figure 3:Bigram Frequency

While there is plenty of positive word pairings **“Love game” and “Fun game**”, I did also notice ads being mentioned several times with the most frequent terms being **“Remove ads”** 91 times**, “Ads support**” 60 times and **“ads forever”** 59 times.

Although it does not appear as frequent I did also see pairings such as “Learning tons”, “Sharp mind” and “School kids” appear 33, 28 and 17 times respectively. Showing that the game is having a great positive impact on some people

Besides just having a look at a table, one of the best ways to visualize bigram pairs are with a spider network.

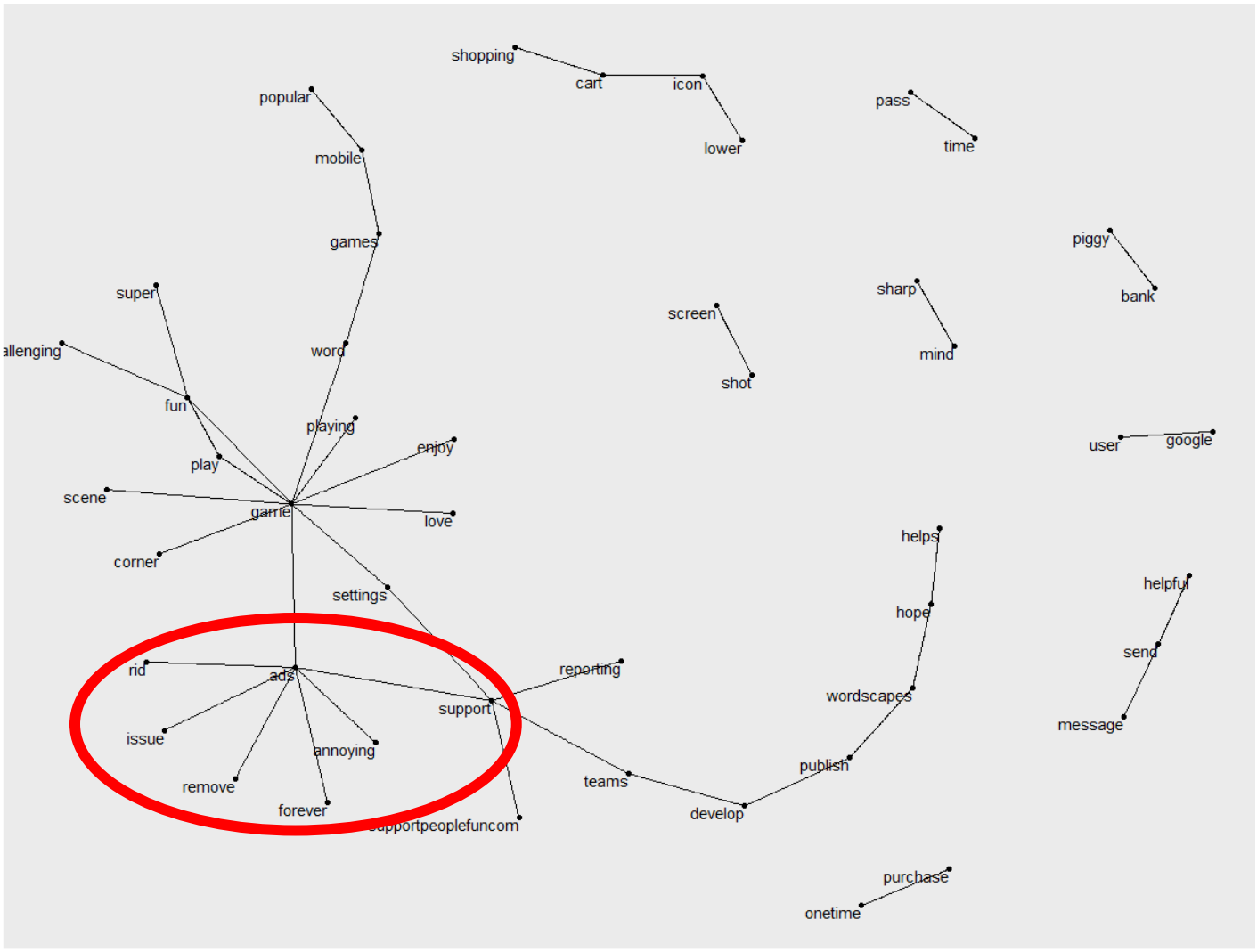

Figure 4: Bigram Network Diagram

Bigram spider networks link words that are highly correlated with each other. Again we notice a common theme of ads being associated with negative words. Yes, we don’t expect everyone to be happy about ads, but we should monitor and ensure that the customers are still having a good gaming experience.

The next stage of my analysis is to carry out a sentiment analysis. I used three sentiment libraries available in R studio:

1. **Bing** - Places words into either positive or negative

2. **NRC** - Places words into eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive)

3. **AFINN** - Words rated for valence with an integer between -5 (negative) +5 (positive)

The AFINN is the best numerical measure of sentiment and it gave an overall mean value of -0.04 which is approximately zero. This leads to a pretty neutral balance between positive and negative sentiment.

	sentiment	method
1	-0.04081633	AFINN

Figure 5:Avg Afinn

We can use the bing library to take an even deeper dive into what words are affecting positive and negative sentiment

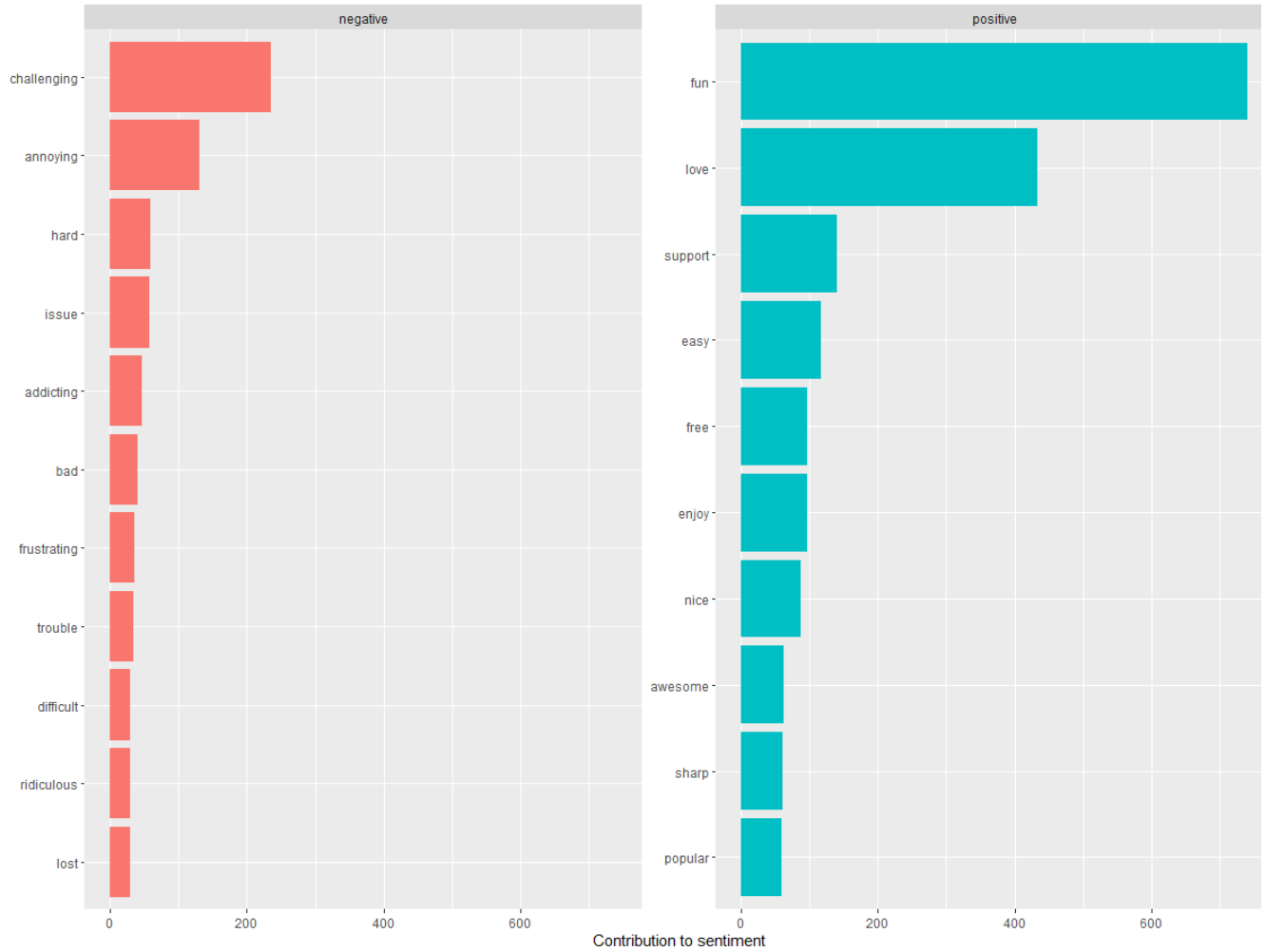

Figure 6: Bing Positive/Negative sentiment

By looking into the positive and negative sentiment we can see that words like “Love” and “Fun” have a strong effect on positive sentiment. On the other hand “Challenging” seems to have the strongest contribution to negative sentiment however the word “Challenging” here is probably not being used in a negative connotation even though the Bing library is picking it up as negative (limitation of Bing), we saw that in Figure 4 that “Challenging” is associated to “Fun”. However, the next biggest contributor to negative sentiment is “Annoying”, we already saw earlier that this is probably referring to the ads (I would investigate the frequency and type of ads users are seeing)

Finally by looking into NRC. I put the 150 most frequent words into a “Pizza Pie” word cloud that shows us the words most linked with the 8 emotions described previously

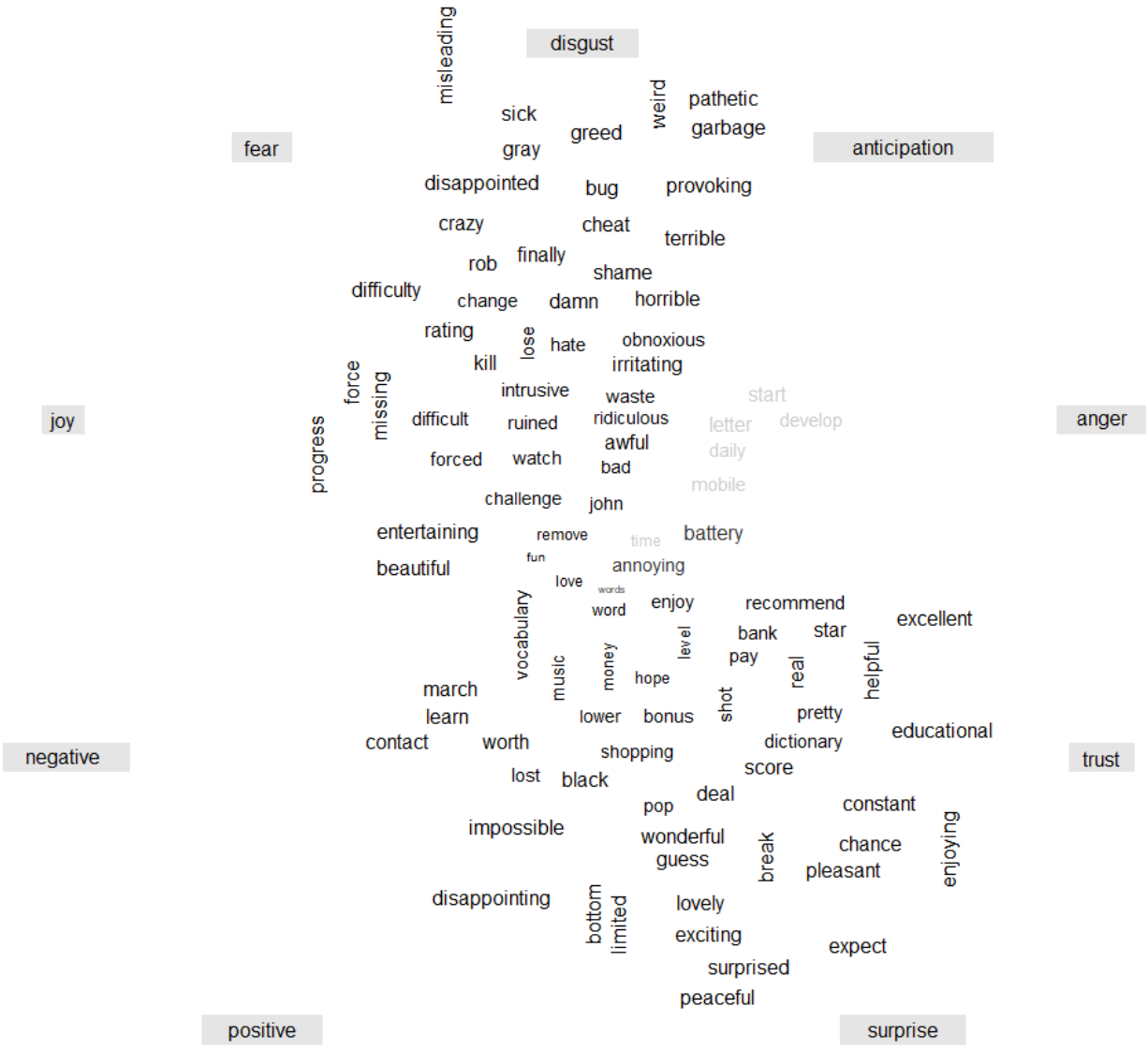

Figure 7:Pizza Pie Word cloud

We notice a lot of words related to surprise and disgust. Although using NRC is valuable since it is the only word library that will put words into emotions, I do find it to be inaccurate sometimes.

Executive Summary:

After scaping over more than 2500 Google play store reviews for the Wordscape game we were able to determine that overall sentiment is neutral (neither positive or negative).

The words being heavily associated with positive sentiment are very frequent such as “Fun” and “Love”. However more interestingly, the words associated with negative sentiment are also being related to the words “Ads” and “Ad” common word pairings include “remove ads”, “ads support” “ads forever”. This requires further investigation into the frequency and type of ads being displayed to users of this game.

NB: As a note, we should also consider the fact that most users of a game never review the game and those who have a bad experience on the app are more likely to leave a review (Qualtrics,2019). This can introduce biases into the data.